Searching the Deep Web

Ms. Greta E. Marlatt

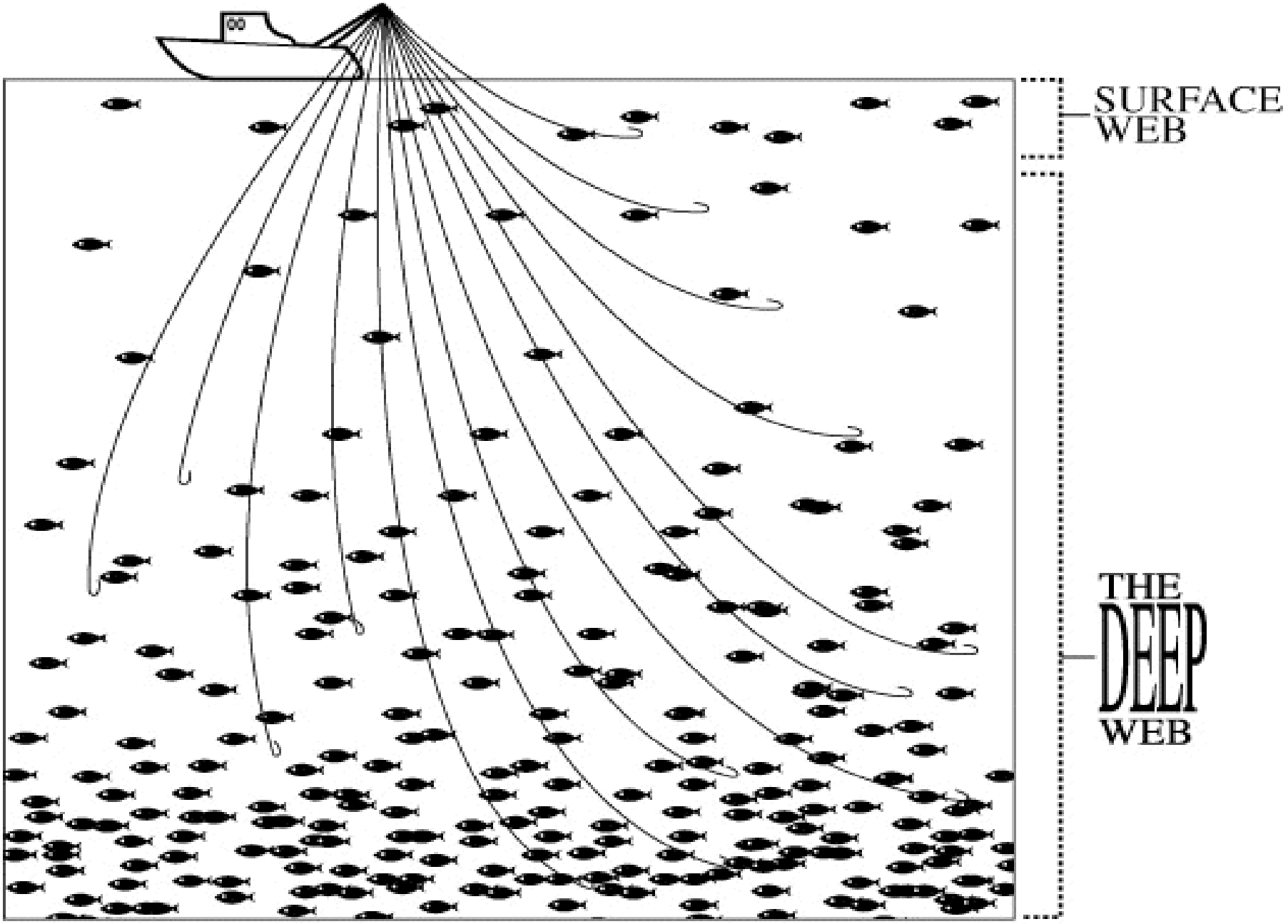
Dudley Knox Library

Naval Postgraduate School

“Deep” or “Invisible” Web

- **What is in it?**
  - “Stuff” that search engines can’t or won’t search
  - Dynamically generated pages [cgi, asp, cfm, URL with ?]
  - Web accessible databases that require information be typed in
  - Password or login required
  - Certain file formats [e.g. PDF, images, streaming media, etc]
  - A lot of the real time data [stock quotes, weather, sports scores, election results etc]
  - Commercial resources with domain or IP limitations
  - Information residing on an Intranet
  - Archives [newspapers]
- **How big is it?**
  - 400-550 times larger than the surface web
  - 7,500 terabytes vs. 19 terabytes

http://www.brightplanet.com/deepcontent/deep_web_faq.asp#DeepWebSize



http://home.lausitz.net/a.krause/tiefe_tiefen.htm

Quick Search
Scientific and Technical Documents

This Quick Search searches all databases checked. Choose this search if you want broad-based search results.
The maximum number of documents returned is 25.

Choose one or more of the following :

- [x] DTIC's Technical Reports Collection
- [x] DTIC's Full-Text Collection
- [ ] DTIC Thesaurus
- [ ] R&D Descriptive Summaries (RDDS)

Note: The more you select, the slower the response time.

Search for : View Database Descriptions

"homeland security" Search Clear Query

Fielded Search Search Help

Spiders can't type

Hint : *Secure STINET gives you access to unlimited and limited distribution citations received by DTIC since 1985 as well as full text reports and Research Summaries (RS) reports from 1965 to present. Find out how to* ***register with DTIC and subscribe****.*

Why Use The Deep Web?

- Quality of content/higher level of authority
- Comprehensiveness
- Focused
- Timeliness
- The material isn't available elsewhere on the Web

When To Use The Deep Web

- Standard search engines aren't working
- A precise answer is needed
- Data or statistics are needed
- High quality or authoritative results are needed
- When timeliness is important
- You know the subject area well
- Looking for collections [images, sounds, manuscripts etc.]
- Reference books online [handbooks, guides, dictionaries, encyclopedias etc.]

Strategies for Searching The Deep Web

- Have the mindset of a hunter, detective or passionate collector
- Use search engines to get to the front door
  - Plane crash database
  - Toxic chemicals database
- Use the site map to see if databases or statistics are mentioned
- Use the site's internal search tool and search for "database"
- Check subject specific discussion groups

What Tools To Use

CALL's Military Domain Search Engine

http://call-search.leavenworth.army.mil/

Searches Multiple (Military related) Sites, including Army, Army field Manuals, Corps of Engineers, Joint Services, Mitre Corp., NATO, Navy, Rand, USAF, USCG and USMC.

Welcome to CALL'S

Military Domain Search Engine

To narrow your search, specify either text or time peri

Type text to search for:

officer retention

Search

Search for content modified during the last:

Use mouse to select one or more desired catalogs. Ctrl key with mouse selects multiple catalogs:

- Army
- Corps_of_Engineers
- Joint-Services
- Mitre
- NATO
- Navy
- Rand
- USAF
- USCG
- USMC

Sort results by:

best matches first

Tips

Search

choose to search one or multiple catalogs

Military Domain Search Engine

Approximately 2.5 million records indexed
Click HERE to see the number of records contained in each catalog.

Tips for Searching

To use this search engine, you may search all catalogs by simply typing your query, and clicking on search. This searches all available catalogs and time periods. You can further define your search by selecting what catalogs you wish to search and define a desired time period when documents were last modified. See below for an explanation.

Content:

This resource is created by a web crawler that periodically indexes portions of the military domain, and CALL does not maintain the content of web pages and documents found by this search capability. Responsibility for content lies with the organization publishing the information to the Internet. ***Query results are created dynamically from indexes created from publicly available on-line and other selected sources. The content and hyperlinks displayed as a result of any search nt by CALL for the information, products or services contained therein.***

ngle word or phrase for your search criteria. The search is not case sensitive, and it considers both e character. You my leave this box blank, provided you specify a time period in the "Content Last ed" box.

Content Last Modified:

- If left blank, this field will not be considered.
- The drop down menu lets you pick between 24 hours, 7 days, 30 days or 365 days.
- By picking one of those time periods, you limit your responses to documents that were last modified within a specified time frame.
- NOTE: The quality of this field is dependent on the system in which the document resides. If the date was improperly recorded, it will be improperly reported in this area.

Catalogs:

- arching.
- By holding down the 'Ctrl' key and clicking on the desired catalog, you can select for deselect multiple catalogs.

http://call-search.leavenworth.army.mil/

What Tools To Use (2)

CompletePlanet -- http://www.completeplanet.com/

Covers over 70,000 searchable databases and specialty Search engines

Government -- http://www.completeplanet.com/Government/index.asp

History -- http://www.completeplanet.com/History/index.asp

Military -- http://www.completeplanet.com/Military/index.asp

http://www.completeplanet.com/

What Tools To Use (3)

First Gov -- http://firstgov.gov/

This is the official gateway to all US Government information searching over 51 million web federal and state web pages.

US government telephone directories and email directories -- http://firstgov.gov/Contact/Directories.shtml

Advanced search -- http://firstgov.gov/fgsearch/index.jsp

FIRSTGOV.gov
The U.S. Government's Official Web Portal

Search | Enter Search Term(s) | Go
in | Federal Only | Advanced

- Federal Only
- Federal & All States
- All States
- One State
- Federal & One State

Home | About Us | Site Map | Help | Español | Other Languages | ...resident Bush

E-mail Us Your Questions

for Citizens | for Businesses and Nonprofits | for Federal Employees | Government Gov

By Organization »
- A-Z Agency Index
- Federal Executive
- Federal Legislative
- Federal Judicial
- Cross-Agency Portals
- State Government
- Local Government
- Tribal Government

Contact Your Government »
- E-mail
- Phone
- In-Person

more

Reference Center »
- Data & Statistics
- Government Forms
- Graphics and Photos
- Laws & Regulations

more

Also of Interest...

the White House
President George W. Bush

egov
Learn about the President's electronic government initiatives

Citizens: Get It Done Online!

E-mail This Page

- Find Government Benefits
- Apply for Government Grants
- Shop Government Auctions
- Apply for Government Jobs
- Change Your Address
- E-file Your Taxes
- Renew Your Driver's License
- Mail in Voter Registration
- Get a Passport Application
- Apply for Social Security
- Check Immigration Case Status
- See Lottery Results
- Contact Elected Officials
- Order Consumer Publications

More Online Services

Information by Topic »

- **Benefits and Grants** — Loans, money, funding, financial aid...
- **Consumer Protection** — Consumer action, publications, credit, complaints, scams...
- **Defense and International** — National defense, foreign relations, immigration, visas...
- **Education and Jobs** — Employment, training, careers, labor, wage, workplace...
- **Environment, Energy and Agriculture** — Farms, food production, natural resources, conservation...
- **Family, Home and Community** — Housing, human services, community development...
- **History, Arts and Culture** — Museums, libraries, genealogy, ethnic, traditions...
- **Money and Taxes** — Unclaimed government money, credit, saving, retirement...
- **Public Safety and Law** — Crime, prison, law enforcement, disasters, emergencies...
- **Reference and General Government** — Acronyms, directories, data, forms, photos, laws, libraries, publications, regulations...
- **Science and Technology** — Biology, engineering, space, innovation, communications...
- **Travel and Recreation** — Transportation, air, train, international, ...

News and Features

- Threat Level Raised
- Grandparents Raising Grandchildren
- Make a Difference – Volunteer
- Terrorism Suspect Photos
- Find Local Farmers Markets
- Gas Prices in Your Area
- Medicare Drug Discount Cards
- Popular Baby Names
- Publication of the Week
- State Travel and Tourism Sites
- Weather
- Past News and Features

- Agency Press Releases
- Free E-mail Newsletters
- FirstGov.gov Press Room

By Audience

- Kids
- Parents
- Seniors
- Military and Veterans
- Americans Abroad
- More Audiences

What Tools To Use (4)

GPO ACCESS – http://www.gpoaccess.gov/index.html

Multi-Database Search -- http://www.gpoaccess.gov/multidb.html

Description of the databases searched above -- http://www.access.gpo.gov/su_docs/aces/dcff001.html

GovSearch search engine [GlobalMaster] -- http://www.globemaster.de/govsearch/index.html

[.mil & .gov only - combines SearchMil.com and SearchGov.com]

Multi-Database Search

About the Databases
Access Methods
Depository Libraries
FDLP Gateways
Specialized Search
Training Manual

Select one OR MORE databases and enter your search below. There are several ways to search across multiple databases from within a scrolling text box. To perform a fielded search of an individual database, go to the specialized search page. Helpful Hints provide instructions for searching the databases. The default database is the current volume of the *Federal Register*.

- Budget of the US, FY2003
- Budget of the US, FY2002
- Budget of the US, FY2001
- Budget of the US, FY2000
- Budget of the US, FY1999
- Budget of the US, FY1998
- Budget of the US, FY1997
- Campaign Reform Hearing
- Congressional Bills, 103d Congress (1993-1994)
- Congressional Bills, 104th Congress (1995-1996)
- Congressional Bills, 105th Congress (1997-1998)
- Congressional Bills, 106th Congress (1999-2000)
- Congressional Bills, 107th Congress (2001-2002)
- Congressional Directory, 105th Congress (1997-1998)
- Congressional Directory, 106th Congress (1999-2000)

Search Terms:

"embassy bombings"

Maximum Records Returned: 40 Default is 40. Maximum is 200.

SUBMIT CLEAR

Phrases must be in quotation marks (""). Word roots can be searched using an asterisk (*) following the word stem. For example: legislat* will retrieve both legislation and legislative. Additional instructions and examples.

http://www.access.gpo.gov/su_docs/multidb.html

What Tools To Use (5)

Scirus -- http://www.scirus.com/srsapp/

This site concentrates on scientific information and covers over 167 million science-specific Web pages. It searches both free and journal sources. Their spiders crawl deeper and do recognize non-text files such as PDF.

Science.gov -- http://science.gov/

This site concentrates on government sites conducting scientific and technical research. It covers authoritative science Web sites and databases of technical reports, journal articles, conference proceedings and other published materials.

SCIRUS
for scientific information only

About Us | Newsroom | Advisory Board | Submit Web Site | Search Tips | Contact Us

Advanced Search Basic Search Advanced Search Search Preferences

chemical — The complete document — All of the words

AND

terrorism — The complete document — All of the words

The complete document
Article title
Journal title
Author(s) name
Author affiliation(s)
Keyword(s)
ISSN
(Part of a) URL

Published between	1930 and 2005
Information types	☑ All ☐ Abstracts ☐ Articles ☐ Books ☐ Company homepages ☐ Conferences ☐ Patents ☐ Preprints ☐ Scientist homepages
File formats	☑ All ☐ HTML ☐ PDF
Content sources	☑ **All Journal sources** ☐ Beilstein on ChemWeb ☐ BioMed Central ☐ MEDLINE on BioMedNet ☐ ScienceDirect ☐ Society for Industrial & App. Mathematics ☑ **All Web sources** ☐ Chemistry Preprint Server ☐ CogPrints ☐ E-Print ArXiv ☐ Computer Science Preprint Server ☐ Mathematics Preprint Server ☐ NASA ☐ US Patent Office ☐ Other
Subject areas	☑ All ☐ Agricultural and Biological Sciences ☐ Astronomy ☐ Life Sciences ☐ Materials Science ☐ Mathematics

Introducing science.gov version 2.0!

Science.gov is a gateway to authoritative selected science information provided by U.S. Government agencies, including research and development results.

Science.gov Search acoutic and mine and d Begin Search Advanced Search

Enables you to search 47 million pages in real time

Explore Selected Science Web Sites by Topic

Agriculture & Food
Food Safety, Gardening, Pesticides, Veterinary Science ...

Applied Science & Technologies
Biotechnology, Electronics, Engineering, Transport ...

Astronomy & Space
Exploration, Planets, Space Technologies ...

Biology & Nature
Animals & Plants, Ecology, Genetics, Pest Control ...

Computers & Communication
Networks, Hardware, Software ...

Earth & Ocean Sciences
Land, Maps, Natural Disasters, Oceans, Weather ...

Energy & Energy Conservation
Energy Use, Fossil Fuel, Solar, Wind ...

Environment & Environmental Quality
Air/Water/Noise Quality, Cleanup, Climate Change ...

Health & Medicine
Disease, Health Care, Nutrition, Mental Health ...

Math, Physics, & Chemistry
Astrophysics, Chemicals, Mathematical Modeling ...

Natural Resources & Conservation
Ecosystems, Energy Resources, Forest Science, Mining ...

Science Education
Homework Help, Teaching Aids, Science Internships ...

Featured Web Sites

Treesearch GPO Access: Science and Technology Drugs@FDA

Search capabilities provided by DOE/OSTI and USGS
Web sites by Topic maintained by NTIS.

What Tools To Use (6)

IncyWincy: The Invisible Web Search Engine

http://www.incywincy.com/

Government -- http://www.incywincy.com/default?p=48669

Military -- http://www.incywincy.com/default?p=371

Infomine Multiple Database Search -- http://infomine.ucr.edu/search.phtml

Infomine might be called an "academic" search engine, since it focuses on scholarly resource collections, electronic journals and books, online library card catalogs, and directories of researchers.

Government Information -- http://infomine.ucr.edu/cgi-bin/search?govpub

What Tools To Use (7)

Invisible Web Directory -- http://invisible-web.net/

This site is a companion to *The Invisible Web: Finding Hidden Internet Resources Search Engines Can't See* by Chris Sherman and Gary Price.

Librarian's Index to the Internet [LII] -- http://lii.org/

Government -- http://lii.org/search/file/government

Military -- http://lii.org/search/file/military

What Tools To Use (8)

Navigating the Military Internet

http://library.nps.navy.mil/home/milinternet.htm

This page, while not a search tool per se, has a lot of good links to all kinds of military information. A related page Where to Find Military Information is also useful since it is organized by topic and leads to many databases containing military information. --

http://library.nps.navy.mil/home/militaryinfo.htm

Questions?

Exercises

- 1) Find the Bureau of Labor Statistics (BLS) Inflation Calculator and discover what the buying power of $100 spend in 1948 would be today
- 2) Find a site that will allow you to see when the sun will be rising and setting in Monterey and Basra (30.49 N, 47.82 E) on June 13, 2008
- 3) Where would you go to find out if McDonalds holds the trademark on the term “Big Mac Attack?”

Exercises

- 4) Where would you go to find a map showing real time earthquake data?
- 5) Where can you go to find USAID overseas loans and grants information broken down by country [hint – limit your search to the .gov domain]
- 6) You're getting ready to deploy Japan, how can you find out what the Yen is currently worth
- 7) You have orders to the Pentagon when your graduate, where can you go to calculate your Basic Housing Allowance [BHA] for Arlington, VA?

Exercise Answers

- 1) BLS Inflation Calculator -- http://www.bls.gov/cpi/
- 2) U.S. Naval Observatory Sun Rise and Sun Set data-- http://aa.usno.navy.mil/data/docs/RS_OneDay.html
- 3) US Patent and Trademark Office Trademark Electronic Search System (TESS) -- http://tess2.uspto.gov/bin/gate.exe?f=tess&state=9bldbv.1.1
- 4) USGS Earthquake Website [with maps of recent activity] -- http://earthquake.usgs.gov/

Exercise Answers (2)

- **5) USAID Overseas Loans and Grants -- http://qesdb.cdie.org/gbk/**
- **6) Federal Reserve Board Foreign Exchange Rates [updated daily] - http://www.federalreserve.gov/releases/h10/update/**
- **7) DOD BAH rates [by year, Duty ZIP CODE and rank -- http://www.dtic.mil/perdiem/bahform.html**